MS39 Compilation of data from published papers and list of knowledge gaps

The main focus of the literature-based data mining activities will be to acquire response functions for the first stage improvement in model parameterisation, of relevance to C3 and C4 modelling. We will initially focus on the effects of ozone, alone and in combination with other pollutants and environmental stressors, on the processes associated with stomatal conductance, photosynthesis and carbon allocation. Later, data collection will focus on those specific aspects of N impacts for which model improvement is required. Participants will search for data based on the vegetation categories: crops, trees, grasslands, wetlands and heath. Data will be collected using a standard template that will facilitate cross vegetation type analysis to look for commonalities in responses. A meta-analysis will also be conducted. We will focus on responses to chronic ozone exposure and realistic N deposition rates.

Here we provide timescales and divisions of work together with criteria for acceptance of data, the key words for searching and a standard format for data reporting and file naming. For each suitable paper found, an Excel template should be complete and uploaded to a dedicated area of the ECLAIRE web-page. Due to copyright issues, we are unfortunately not able to upload the PDFs of the papers to a central location. However, the abstract, relevant figures and tables will be copied and pasted into the Excel file, should the user of the data need to refer back to the originals.

It will then be possible to search the database on ECLAIRE for specific information/data in the papers, to be outputted as an Excel spreadsheet for each search. As indicated below, we are looking for specific papers that fit the criteria. All searches should be retained by the data miner as word documents with the reference and abstracts in, in case at a later stage, we need to widen the data acceptance criteria for specific processes.

A1.1 Expected endpoints for data mining

In the first phase, data will be used to form dose-response relationships and time-courses for the effects of ozone, alone or in combination with nitrogen, ammonia, CO2, drought or warming on the following:

Plant process parameters:

- A-C_i curves (net CO_2 assimilation rate, A, versus calculated sub-stomatal [or chloroplast] CO_2 concentration, C_i), used to calculate: V_{cmax} (Maximum carboxylation velocity), J_{max} (maximum rate of electron transport)
- Rubisco limited photosynthesis, RuBP-limited photosynthesis, TPU limited photosynthesis
- R_d (dark respiration rate)
- Protein turnover rate

- A-Q curves (net CO_2 assimilation rate, A, versus photon flux density, Q) to calculate A_{sat} (light saturated photosynthesis rate) and Φ (maximum quantum yield of photosynthesis). The temperature parameters of these curves would be useful.

- Net ecosystem production (the difference between gross primary production and total respiration of the ecosystem)

- g_{max} (maximum stomatal conductance) and g_{min} (minimum stomatal conductance);

- Leaf starch content (as an indication of phloem loading)

  DO_3SE model parameterisations (see LRTAP Convention, 2010, Modelling and mapping Manual, Chapter 3, for details, downloadable from http://icpvegetation.ceh.ac.uk/manuals/index.html). Please copy and paste the table into the Excel file.

Plant growth related parameters

[**Note:** we are not collecting data on yield or growth end-points such as total biomass, other than those described as this data collation is for improvement of process/physiology within the models]

- biomass partitioning into the roots, stems, reproductive parts including seeds, leaves

- Leaf senescence (including chlorophyll content and (non-structural) N content)

- C:N ratio and C and N content of leaves, roots, stems, and reproductive parts

- Leaf area index

- Specific leaf area

- Fine root turnover

Soil process related parameters

- Soil dissolved organic carbon content

- Soil microbial activity

A1.2 Work-plan and who will do what

A1.2.1 Timescales for data collection, analysis and reporting

Please see GANTT chart in **Table A1.1**. The data collection and analysis will occur in four phases:

Phase 1: Data collection should focus on effects of ozone, alone and in combination with other pollutants and environmental stresses on plant processes related to photosynthesis

Phase 2: Analysis of phase 1 data

Phase 3: Data collection for selected processes for N (details to be decided later)

Phase 4: Analysis of phase 3 data

Table A1.1: GANTT Chart describing timing of activities for the literature-based data mining

WP9/14 GANTT chart	2012					2013										
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov
ECLAIRE month	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Finalize methodology																
Progress report																
Progress meeting			ECLAIRE meeting			ICP Veg meeting					group meeting?				ECLAIRE meeting	
Phase 1 data collection (O3 & interactions, plant																
Phase 2- analysis of ph.1 data																
Phase 3 data collection (N & interactions, selected data)																
Phase 4- analysis of ph.3 data																
Completion of data collection and analysis																

A1.2.2 Division of work between research groups

Staff time available for the datamining (units: person-months) is:

> **C3, WP9**: CEHB (7); CEHE (2); DTU (2); UGOT (3); IVL (1.8); SEIY (possibly some left over from ecosystem-scale data collation)
>
> **C9, WP14**: Coordinated by JRC with other partners (15).

See **Table A1.2** for who will do what.

A1.3 Literature-based data mining method

Important: For each datapoint please record: mean, standard deviation and the number of replicates (n). This will facilitate derivation of dose-response relationships and meta-analysis.

A1.3.1 Selection of papers

Publication status:

- Studies should be published in peer-review scientific journals or alternatively unpublished by well recognized scientists.
- Studies should be from Europe, USA or Canada.

Table A1.2: Division of literature based data mining between groups, including species choices for trees and crops

Veg class	Sp. or veg type and/or analysis tasks	Lead	Institute	No of person months
Trees[1]	Trees dose-response and meta-analysis	Per Erik Karlsson	IVL	1.8
	All species listed in footnote 1	Alessandro Cescatti	JRC	10
Heath & wetlands	Dry heath	Claus Beier	DTU	2
	Wetlands	Lucy Sheppard	CEHE	1
Grasslands	Northern/central Europe upland and lowland conservation grassland; Mediterranean grassland	Alessandro Cescatti	WP14	6
Crops[2]	wheat, barley, potato, bean, soybean; Crops dose-response and meta-analysis	Hakan Pleijel	UGOT	3
	maize, tomato	Lucy Sheppard	CEHE	1
Coordination, reports, meta-analysis etc	Dose-response and Meta-analysis for grassland, heath and wetland species, and cross-species/veg types data analysis	Gina Mills	CEHB	7

[1] Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), European Silver fir (*Abies alba*), Aleppo pine (*Pinus halepensis*), Maritime pine (*Pinus pinaster*), Larch (*Larix sp*), Beech (*Fagus sylvatica*); silver birch (*Betula pendula*);downy birch (betula pubescens), ash (*Fraxinus excelcior*), Poplar sp, *Salix sp.,* Oak (*Quercus robur*, *Q. petreae*), English, sessile, Holm (*Quercus ilex*)

[2] wheat (*Triticum aestivum*); Maize (*Zea mays;* Barley *(Hordeum vulgare);* Soybean *(Glycine max);* Bean (*Phaseolus vulgaris*)*);* Potato (*Solanum tuberosum);* Tomato (*Solanum lycopersicum)*

Choice of exposure system

Data from experiments conducted using the following exposure systems are acceptable:

- Open top chambers
- Field release systems
- Field manipulation sites (e.g. with added N)
- Ambient air wet N applications
- Solardomes
- Field epidemiological correlation studies
- Field gradient studies

Note: Please record data from papers describing controlled environment room experiments (where climatic conditions are close to ambient and other criteria related to exposure time are met). However, such data will not be included in initial analysis but may be used later, if needed.

Method of growing plants

- Ideally naturally growing or planted in the open field.
- Data from pot-based experiments are accepted providing the pot size provides adequate root growth space (i.e. dependant on species, but with pot size exceeding at least 5 litres).
- For trees – the minimum height at the end of the exposure should exceed 50cm and/or the dry weight should exceed 200g.

Length of exposure/treatment period

- For crops and grasslands: experiment should last for a growing season or at least three months
- For trees and other perennial vegetation: Data should only be included for experiments that run for at least one growing season.

Treatment ranges accepted

Ozone: Seasonal/exposure period daylight (or 7, 8 or 12h) hour mean ozone concentration does not exceed 100 ppb; if a paper has some treatments below 100ppb and some above, record all of the data but that from treatments above 100 ppb may not be used in the meta-analysis.

Nitrogen: highest application rate should not exceed 80 kg N ha^{-1}; if a paper has some treatments below 80 kg N ha^{-1} and some above, record all of the data in the paper but that from treatments above 80 kg N ha^{-1} is unlikely to be used in the meta-analysis.

Ammonia: highest concentration should not exceed 50 µg m^{-3}.

CO2: up to double ambient.

1.3.2 Key words for literature search

Species name (English and Latin) or vegetation type, with:

- Photosynthesis [interested in RuBisco, V_{cmax}, J_{max}, Asat, quantum yield, mesophyll conductance, Rubisco limited photosynthesis, RuBP-limited photosynthesis, TPU limited photosynthesis, light saturated photosynthesis rate, maximum quantum yield]
- Net Ecosystem Exchange or Gross Primary Productivity
- respiration
- stomatal conductance
- carbon allocation
- biomass partitioning
- phloem loading
- protein turnover
- nitrogen content
- leaf starch content
- senescence
- leaf area index
- specific leaf area
- Dissolved organic carbon [in soil]
- soil microbial activity
- fine root turnover
- Soil organic matter
- Soil respiration
- Soil microbial activity

and:

- ozone, alone and in combination with
  - nitrogen
  - ammonia
  - CO2
  - drought
  - warming

A1.3.3 Data recording

Important note: the Excel template file has to be stored as an Excel 97-2003 workbook, with the file ending of ".xls". This is essential for uploading to the Excel database.

Please enter the data for each paper into a separate provided Excel template file. The template has separate worksheets for the paper metadata, pollutant and other treatments, vegetation details, plant process data, plant growth and biomass data, and soil process data. There is also space to copy and paste the abstract and the relevant figures and tables from the paper, with the latter best copied using the snap-shot facility of Adobe Acrobat X – Pro.

For each datapoint please record: mean, standard deviation and the number of replicates (n). This will facilitate cross-species and cross-vegetation type derivation of dose-response relationships and meta-analysis.

Enter the data for each paper into the template file provided. The keywords in the first spreadsheet will facilitate data searching within an Oracle-based ECLAIRE database, allowing all of those files to be found that include data for combinations of searched for terms. By using the same template for all files, the data can be combined and analysed as needed. For each file, please complete all worksheets as appropriate for the paper. All data within the file will be associated with a unique

identifier generated by the software. This identifier will be generated in cell B15 of the "Data Mining Metadata" worksheet and should be used as the filename. It has the following name format: First author name_year published_start page number, for example: Mills2010123 for a paper with Mills as first author, published in 2010 and start page number of 123.